DOCUMENT RESUME

ED 244 563 HE 017 266

AUTHOR Yost, Michael
TITLE Extracting Data From Integrated Student Information Systems.
PUB DATE Dec 83
NOTE 17p.; In: Information Resources and the Individual. Proceedings of the CAUSE National Conference (San Francisco, CA, December 11-14, 1983). Boulder, CO, CAUSE, 1984, p. 297-310. For the complete proceedings, see HE 017 245.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Administration; *Computer Oriented Programs; Data Analysis; *Data Processing; *Documentation; Higher Education; *Information Retrieval; Institutional Research; *Management Information Systems; Recordkeeping; Scheduling; *Student Records
IDENTIFIERS Faculty Records

ABSTRACT

An integrated student information system (ISIS) at Trinity University (Texas) is described with attention to how data are entered and how data are extracted for purposes of institutional research. The structure of the original ISIS files, the extract files, and the Statistical Package for the Social Sciences (SPSS) files is examined. An outline is provided of the methods used for transferring data from one file to another. Attention is also directed to methods used to develop and maintain the files in an interaction course management system, along with methods to analyze data using SPSS. There are about 1,500 variables in the ISIS, which contains all student academic and personal data, all faculty data that deal with courses and teaching, all course data that deal with students, faculty, class meeting times and days, enrollment, buildings, and the college calendar and catalog. Three programs that were designed to extract data from the ISIS to conduct institutional research and to develop ad hoc reports are: the Prospect/Applicant Extract, which extracts data on prospective students and/or persons who applied for admission; the Enrolled Student Extract, which extracts data on enrolled students, and the Course/Faculty Load Extract, which extracts data for all courses taught in a semester. Outlines of the content of the enrolled student extract and the course/faculty load extract are included. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED244563

Extracting Data From

Integrated Student Information Systems

Michael Yost, Ph.D.
Director, Institutional Research
Trinity University
San Antonio, Texas 78284

ABSTRACT

This paper describes an ISIS, how data are put into it, and how data are extracted from it for purposes of institutional research. The structure of the original ISIS files, the extract files, and theS PSS SPSS files is presented and the methods used for transferring data from one file to another are outlines. The methods used to develop and maintain the files in an interactive CMS computer environment are also given. Finally, the methods used to analyze the data using SPSS are presented.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CAUSE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 017 266

Paper presented at the
CAUSE National Conference
(San Francisco, CA, Dec. 11-14, 1983)

Extracting Data From Integrated Student Information Systems

Introduction

Two years ago, we at Trinity University implemented an integrated student information system (ISIS). Figure 1 is a graphic representation of the way in which students are processed as they enter and progress through our institution. All of our data are maintained in our IBM 370-3031 computer. Each of the academic offices on campus contains computer terminals which permit persons who work there to access the database and to update or modify selected data elements in the database. When a prospective student applies for admission, a file for this individual is "opened" in the system and all of the application data are entered. Later, when the individual enrolls, he/she processes through the offices on campus. The file which was "opened" in the admissions office is available for expansion to the personnel in each of the other offices. As a student is processed through each office (registered, billed, housed, etc.), his/her data are entered into a single master file. Conducting institutional research from this database, where all of the data on students are held in the same place, is far different from conducting institutional research using a number of different data files which traditionally are separately maintained in a number of different offices. In our ISIS, the data for each student are loaded in the master file. The subsections of the master file for each student are cross-referenced. This arrangement makes it a relatively simple process to acquire data needed to conduct institutional research.

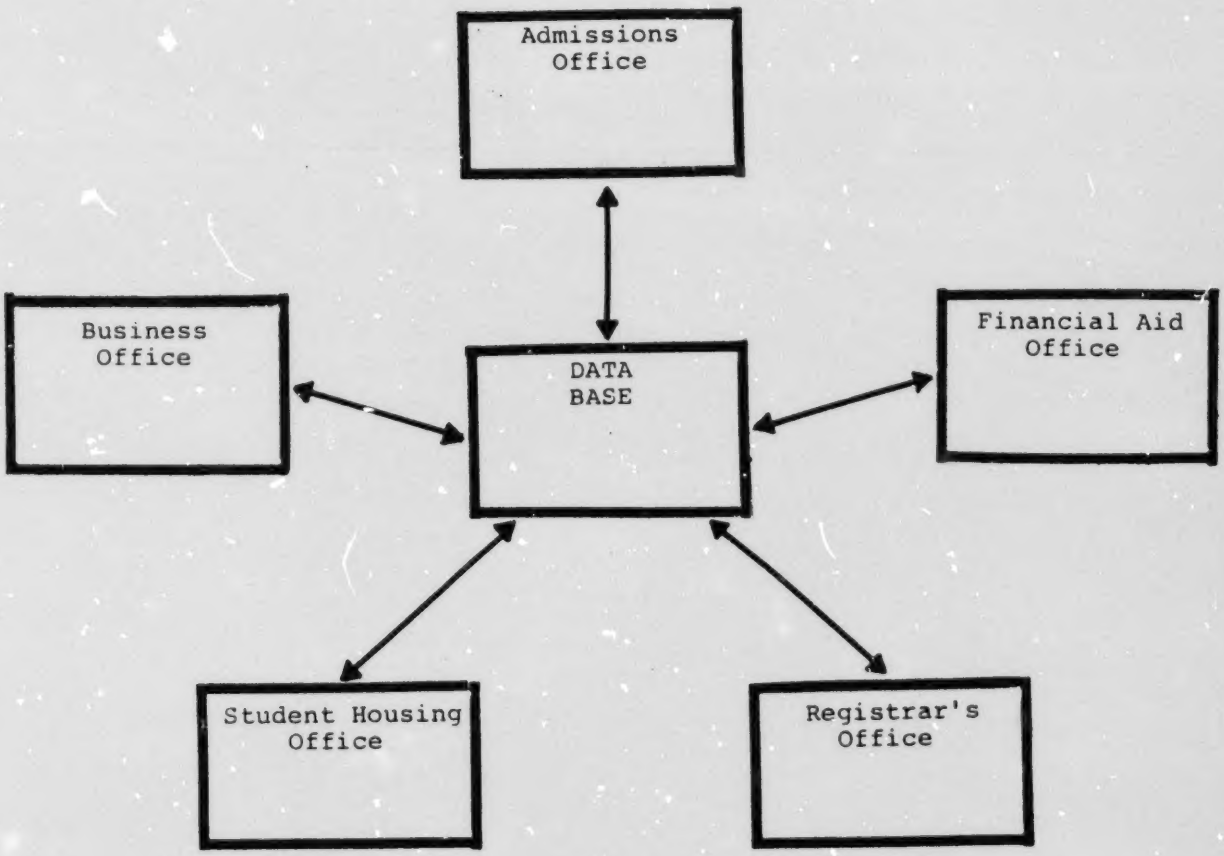
Data and File Structure

To understand how institutional research can be conducted using an ISIS, one must first understand which data elements are input, how the data are maintained, and how the files are structured. There are approximately 1500 variables in our ISIS. Many of the variables can have two or more entries for a given student, faculty member, or class. As a result, two students, faculty members, or classes seldom have the same number of variables in their files.

The data are input to the ISIS through 65 on-line terminal screens into 54 separate subfiles (See Figure 2).

Table 1

Record-Keeping Under An
Integrated Student Information System



In addition, some variables are maintained by the system and others are maintained through the use of batch programs which "roll" the data into particular fields. For example, a student's date of birth would be input on an on-line basis through a terminal; a bill for room, board, and tuition would be system-maintained on an on-line basis and automatically calculated; and the number of credit hours generated by a professor in a given course on the census day of a semester is maintained through a batch program which calculates the correct value and "rolls" it into the system. The batch or "roll" programs are ones which are run at specific times during a semester to update files which must all be updated at the same time in a semester or academic year.

The ISIS contains all student academic and personal data, all faculty data which deal with courses and teaching, all course data which deal with students, faculty, class meeting times, class meeting days, enrollment, buildings, the University calendar, catalog, and class schedules.

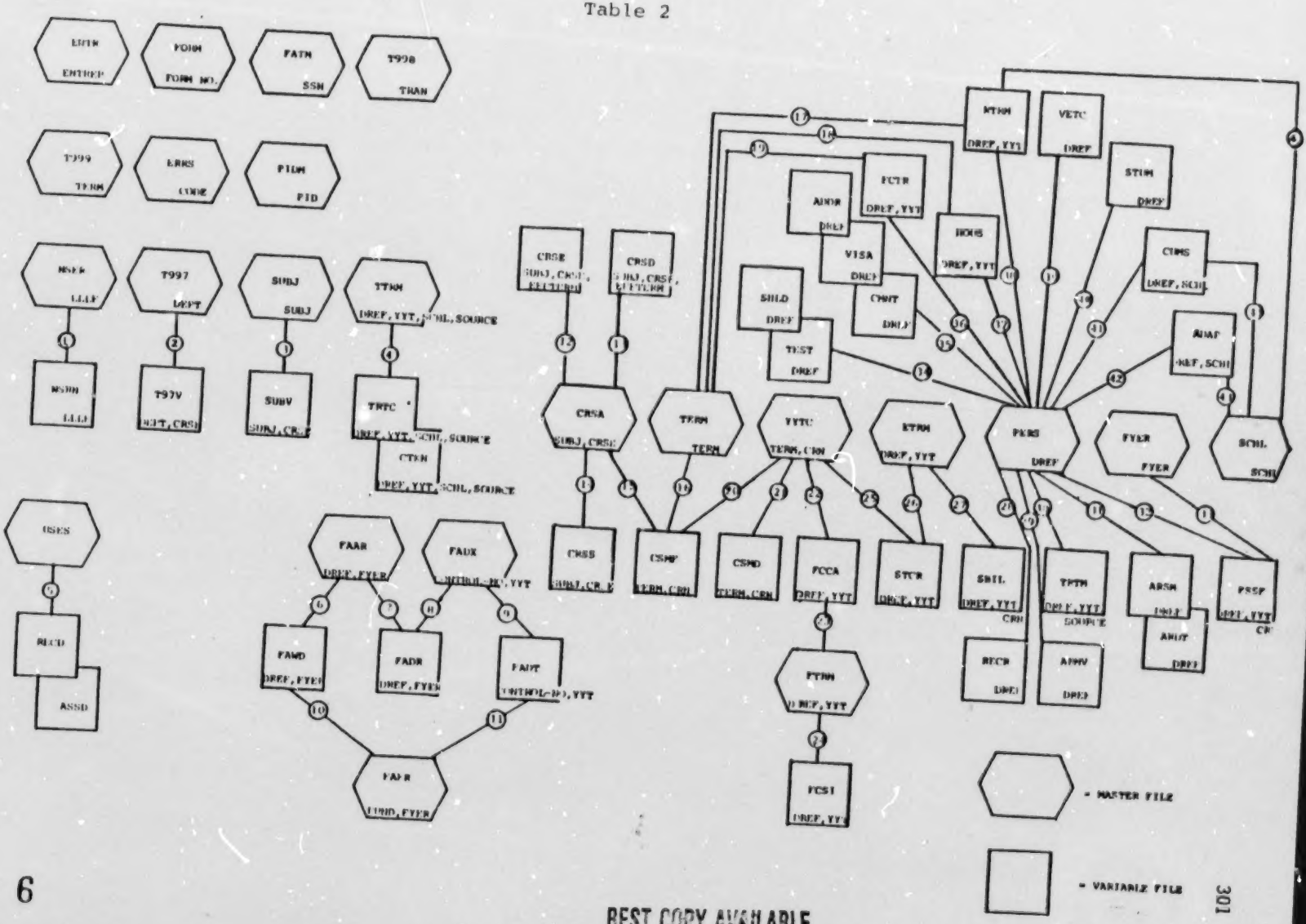
The majority of the ISIS files are BDAM (Basic Data Access Method) files. Several of the files are VSAM (Virtual System Access Method) files. TOTAL is used to manage the files in our ISIS.

Accessing Data

It is a relatively simple and straightforward process to access data on an individual student, faculty member, course, etc. in our ISIS. If a person has the proper passwords, he can sign on to the system at a terminal, and by using the appropriate codes he can call up the data that he wants on one or more terminal screens. If a printed copy of the information on a screen is required, there are several sites on campus where the information that has been brought up on a screen can be printed.

Along with the on-line programs which are used to enter and retrieve data for individual students, there are a series of batch programs which are used to develop standard reports. These reports are those which are periodically required by the individuals in the different offices on campus. The batch programs which prepare these reports were developed at the time our ISIS was designed and installed. The reports which these batch programs develop are similar to the ones which business and academic offices on campuses commonly use. The batch programs which prepare the reports are very functional for their intended use, but they are not very flexible

Table 2



BEST COPY AVAILABLE

and do not produce data in a form which is useful for institutional research purposes. Also, the batch programs do not permit the development of ad-hoc reports which are often needed.

While the accessing of data in ISIS for a particular observation is relatively simple, obtaining aggregated data is a very difficult and complex task. When we purchased the license to use TOTAL as our database manager, we had an opportunity to purchase the license to a second portion of TOTAL which would have permitted us to access and aggregate data directly from the files. This second option costs approximately $40,000.00, and we decided not to purchase the license for it.

We, therefore, designed and developed three programs to extract the data from the ISIS to conduct institutional research and to develop ad-hoc reports. The first program extracts data from the ISIS on prospective students and/or persons who made application for admission (Prospect/Applicant Extract). The second program extracts data on graduate and/or undergraduate students who are enrolled (Enrolled Student Extract). The third program extracts data for all of the courses taught in a semester (Course/Faculty Load Extract).

An outline of the content of the Enrolled Student Extract is contained in Table 3. There are eight major divisions of the content of this extract. The first, or personal information portion, contains information on the sex, age, ethnicity, religion, etc. of each student. The second, or demographic portion of the extract, contains information about each student's parents and their occupations. Also, this section contains each student's complete home address as well as his/her address while enrolled at school.

The third, or academic history portion of the extract, contains information as to which high school the student attended, his high school grade-point average (GPA), and his/her high school rank. For students who are transfer students, the college previously attended, the GPA obtained at that college, and the number of credit hours obtained are also contained in the extract.

The fourth, or academic aptitude portion of the extract, contains the standard aptitude and/or achievement scores of each student. The ISIS itself permits the entry of any number of each of the types of standard scores and subtest scores. In instances in which a student has taken a particular test more than once, the extract program uses the highest score which the student has ever obtained on that test.

Table 3

Content of Enrolled Student Extract

1. Personal Information

2. Demographic Information
   - Home
   - School

3. Academic History
   - High School
   - Previous College(s)

4. Academic Aptitude
   - S.A.T.
   - A.C.T.
   - T.O.E.F.L.
   - G.R.E.

5. Recruiting Activity

6. Academics
   - Major(s)
   - Performance
   - Standing
   - Degree(s)

7. Financial Awards
   - Grants
   - Scholarships

8. Business Office Activities

Table 4

Content of the Course/Faculty Load Extract

1. Instructors
   - Name
   - Identification Number
   - Home department
   - Home division

2. Course
   - Identification number
   - Department identification number
   - Divisional identification number
   - Credit hours
   - Type (lecture, lab, etc)
   - Level (Graduate/undergraduate)
   - HEGIS number
   - Number of students ever enrolled
   - Number of students enrolled on the census day
   - Credit hours generated
   - Meeting days
   - Meetind hours
   - Meeting building
   - Meeting room

The fifth, or recruiting activity portion of the extract, contains information on the number of times a student and recruiter interacted, who the recruiter was, the types of information and/or materials which were given to a student, and the dates that the student and recruiter interacted. Basically, the entire recruiting history of a student can be reconstructed from the data contained in the extract.

In many ways, the sixth, or academic portion of the extract, is the most important. This portion of the extract contains the date which a student's application was received, the term which he/she entered college, the major which the student thought he would pursue when he made application for admission, the actual or the declared major of study, credit hours attempted, credit hours completed, advance placement credit hours, credit hours obtained by examination, transfer credit hours, degrees received, and the number of credit hours which a student has registered for in each semester of attendance. These last data elements (registered credit hours for each semester) have proven to be extremely useful in conducting institutional research studies. It is through the use of these values that full and part-time students are identified, that full time equivalent students are calculated, that a "head-count" is obtained, that the full-time equivalent of part-time students is obtained, that students who "sit out" a semester are identified, that students who do and do not preregister for a semester are identified, and that students who permanently drop out of school are identified.

The seventh, or financial awards portion of the extract, contains grants and scholarship information on each student. At Trinity, as at most schools, a student may receive a grant or scholarship from one or more of several different funding sources. The ISIS is capable of accounting for as many different grants and/or scholarships as a student may be given. Since the extract is an abbreviated version of the ISIS, not all of the information contained in the ISIS can be incorporated into the extract. We have found that by including the first five funding sources for each student, we have complete data on approximately 98 percent of students. We feel that the trade-off of space conserved in not including a sixth or seventh funding source for each student is worth not having complete information in the extract on two percent of the students. There are three pieces of information associated with each source of funding. The first is the identification number of the source of funding. The second is the

status of the award (offered to the student, accepted by the student, rejected by the student, disbursed to the student, cancelled). The third is the number of dollars awarded to the student from the funding source. To reduce the total number of data elements which must be used in any analysis, the first two data elements (identification number and status) are formatted into a single data element. The first three digits of the new format account for the funding source and the last two digits indicate the status of that funding source for a student. Using this formatting approach has reduced the number of data elements in this portion of the extract from fifteen to ten, and has made analyses of the data much simpler without losing any of the information contained in the original data.

The eighth, or business office activities portion of the extract, contains information about each student's transactions at the business office. When our ISIS was developed, a unique set of receipting codes was developed. Each time a student completes a financial transaction in the business office, a receipt code is entered into the ISIS along with the dollar amount of the transaction. From an institutional research point of view, the transaction code is just as important as the dollar amount. Examples of several of the more frequently used transaction codes are: RDEP or room deposit; DDDP or dorm damage deposit; and TDEP or tuition deposit. Along with the transaction code, there is a one-digit numeric code which is affixed to the end of each of these codes to indicate whether a student is making a payment to the institution or whether the institution is refunding the money to the student. The extract contains the first five transaction codes for each student in each semester. Using the first five codes permits an analysis of the business office information for all students through the first month of each semester and for 95 percent of the students for the entire semester. That is, as the semester progresses, approximately 5 percent of the enrolled students have more than five business office transactions. In this case, as it was with financial awards, the conservation of space in the extract using this approach is worth not having some of the information available for analysis.

Each student's personal, demographic, academic history, academic aptitude, and recruiting activity data are input to ISIS at the time he/she enters Trinity. Very few changes are made to these data after they are entered. For the other data elements, however, new data are entered as they become available. Housing, academic, and the business office data are all maintained on a semester basis.

A person is considered to be a prospective student if he/she sends a letter of inquiry to us or if he/she fills out an inquiry card and sends it to us requesting information about our institution. As a result of the inquiry, we typically obtain some personal, demographic, academic, and/or aptitude information on a person whom we consider to be a prospective student. This information is usually self-reported, and is recognized as being self-reported and being subject to error and/or bias. A person who is classified as an applicant is one who has completed an application form, has paid an application fee (or had it waived), and has had official copies of materials forwarded to us. The data in the records of a student applicant are much more complete and reliable than the data in the records of a prospective student.

The data in the Prospect/Applicant Extract is identical to the data contained in the first five portions of the Enrolled Student Extract. Much of the computer program which is used for the first five portions of the Enrolled Student Extract is also used in the Prospect/Applicant Extract. The use of essentially the same COBOL code in two different programs facilitated the development of these two programs.

The content and format of the Enrolled Student and Prospect/Applicant Extracts are much different from those of the Course/Faculty Load Extract. The first two extracts produce a data matrix with rows which represent students (enrolled, applicant, or prospective) and columns which represent variables. The Course/Faculty Load Extract produces a data matrix with rows which represent courses (or sections of courses) and columns which represent variables.

The Course/Faculty Load Extract contains two major parts (see Table 4). The first, or instructor information part, contains the information about the person(s) responsible for teaching the course. This includes the instructor's name, and his/her home department, and division. The second, or course information part of the extract, contains identification information about the course, the department offering the course, enrollment in the course (section), where the course meets, and when the course meets.

The three extract programs were written in COBOL by a competent, experienced computer programmer. The programs were not an easy programming task. Each program took approximately two weeks to write. The main difficulty in writing the extract programs was not in the development of the code

itself, but in determining which data elements were required, which file a desired element was located in, and in developing the link-paths needed to be able to access the data.

The validation of the extract programs and the data which they extract from the ISIS database is an important and time-consuming part of the process of developing an institutional research database. Fortunately, we maintain a test system with a test database which contains about five percent as much data as does our "live" system. Initially, the extract programs were run using the test database as the files from which to extract data. The data which were extracted from the test data files were then compared to the known content of these files. Finally, the extract programs were run using the "live" database files. The results of these runs were spot-checked and the number of records extracted were compared to the number of records which several of the batch programs indicated were present within the system.

Using Extract Programs

Once the extract programs were developed and validated, a schedule was developed for their use. The Course/Faculty Load Extract extracts data which is generated in ISIS by a batch program which "rolls" census day (12th class day) enrollment data into particular data elements. The extract is, therefore, scheduled to be run the day following the "roll". This extract is designed to be run for a designated semester. When the program is submitted, it extracts data for the semester which is designated on the control card for the program. Our ISIS is designed so that historical data can be kept in the database. We are now in our eighth semester of the use of our ISIS and the Course/Faculty Load Extract Program can be used to extract data for the current or any of the previous seven semesters. A copy of the Course/Faculty Load Extract data for each of the eight semesters is now kept in our vault. Should we now decide to go back into our ISIS to extract the data using a different approach, we could.

The Prospect/Applicant Extract is run several times each semester to extract data on prospective students. Trinity University is an institution which is very interested in tracking the receipt of applications and the quality of the persons who make application. For 40 weeks of each year the Prospect/Applicant Extract is run to extract data on applicants. The extract is run at the close of the work day each Friday, and the data on the quality and the number of applicants is published on Monday.

The Prospect/Applicant Extract program is a flexible one. The program permits data to be extracted for graduate and/or undergraduate students. Also, the program is designed to extract data for those persons who have applied for admission in one or any of a number of semesters. This program can be used to extract data on persons who have applied for admission in past semesters. The capability of being able to extract data from the current and from the past year(s) provides a great deal of research capability.

The Enrolled Student Extract program is just as flexible as are the other two extract programs. This program will permit data to be extracted for graduate and/or undergraduate students who are/were enrolled in any semester. Also, the number of credit hours for which a student was/is/will be (preregistered) registered can be determined. At present, we can only extract registered credit hours for nine semesters (8 past, 1 future) because that is all we presently have in our system. The extract program is designed so that it will eventually be able to retrieve data for up to 20 semesters for each student.

When data are extracted for semesters prior to the current one, the cumulative data elements in the extract reflect values for the current semester and not for the semester for which the data are extracted. Most of these differences can be adjusted by subtraction prior to an analysis of the data.

The three extract programs provide a great deal of research power to an investigator. With the structure of the ISIS, the researcher no longer needs to go from office to office gathering data but can, instead, obtain integrated data from a single source. Also, the data have already been validated by the individuals who put it into the system. Finally, longitudinal data are readily available and easily obtained.

Using the Extracted Data

Each of the extract programs produces a magnetic tape which contains the data that were extracted. Before beginning the data analysis, the person responsible for conducting institutional research must understand the data that he/she has on the magnetic tape. The data produced by each of the extracts can be viewed as being a series of rows and columns. These rows and columns constitute a data matrix. For two of the extracts, the matrices have rows which represent students (or prospective students), and columns which represent

variables. In analyzing the data, an investigator examines one or more of the variables (columns) for all or for some subset of the students (rows). This means that an investigator must have a computer which will permit him/her to examine variables and to select all or some subset of the students. To "examine" variables implies the use of one or more of several investigative techniques. First, examination may involve simply describing the distribution of responses on a particular variable. This could be as simple as calculating the measures of central tendency and dispersio and producing a simple histogram. Second, examination may involve producing a cross-tabulation of two or more variables, producing a breakdown of one variable given one or more other variables, or producing a coefficient which indicates the strength or the amount of relationship between two or more variables.

The process of examining variables is typically done in combination with choosing some subset of the students on whom the data were gathered. For example, an investigator might be interested in comparing the academic performance and aptitude of male and female students. To do this, he/she could select to compare the GPA and SAT scores of the male and female students. To complete the analysis, the investigator needs to be able to select out the scores of all the male students and to do the necessary calculations on their variables (scores). Once this is completed, he/she must then do the same thing for the scores of the female students. It is combinations of these selection and analysis processes that an institutional researcher carries out as he/she does his/her work.

Conceptualizing a set of data as a matrix can facilitate the process of conducting research since this is the way in which most computer programs carry out their work. To date, there are no small (micro) computers which are capable of analyzing in a reasonable amount of time the data produced by the extract program. A large main frame is therefore required to read the tape and analyze the data.

There are a number of commercially available software packages which can be used to analyze data on a large main frame. The two most popular ones are SPSS and SAS. Neither of these statistical packages is very "user friendly", but both of them contain commands which are very "powerful". Power in this instance means that there are statements or commands in each of these statistical packages that an investigator can use which make the computer do a lot of work for him/her and do not require that the investigator do a great deal of work or have a great deal of knowledge of computers in order to get the computer to accomplish the work.

Using SPSS, an investigator can use one or more SELECT IF statements to identify and select almost any subset of a student body. These statements are very easy to learn and to use. Once the subset of the student body is identified and selected using SELECT IF statements, then statements such as FREQUENCIES, CROSSTABS, BREAKDOWN, CORRELATION, etc can be used to analyze the data. An institutional researcher who is new to computers may need some assistance in formatting data obtained by an extract program and in the initial preparation of a data deck. However, after having obtained some initial assistance, there is no reason why he/she cannot be conducting analyses on his/her own in a few days.

It is difficult to envision an effective director of institutional research who does not have a first- hand knowledge of the data which are collected on a campus, how the data are collected, how the data are formatted, and how the data can be analyzed. Even if this individual does not conduct his/her own analyses, he should have a knowledge of the characteristics of the data in order to effectively direct those individuals who do the work for him/her. As extracted data become more common, computer systems become more available, and institutional researchers become more knowledgeable, institutional researchers will be able to make even larger contributions to the administration of their institutions.